THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 59 | September 1980 | Number 7

The Ultimate Capacity of Frequency-Reuse Communication Satellites

By A. S. ACAMPORA

(Manuscript received April 10, 1979)

Satellite system concepts that employ multiple spot beams with frequency reuse increase in complexity as the transmission capacity is increased and efficiently allocated throughout a service region much wider than the width of a single spot beam. A fundamental question is the ultimate capacity limit of a satellite system constrained in radiated power, bandwidth, and the aperture dimensions of the satellite antenna. Because the satellite system integrates multiple users, the capacity limit is expressed as a region over the space of transmission rates $\mathcal{R}_1, \cdots, \mathcal{R}_G$ *into the G ground terminals. Any particular set of rates contained in this region can be transmitted with arbitrarily high bit error-rate performance. In this paper, we find inner and outer bounds on the capacity region and show them to be dependent upon the geographical distances between ground stations and the aperture dimensions of the satellite, since these control the effectiveness of frequency reuse. The outer bound is all-inclusive and no time-division multiple access, frequency-division multiple access, spread-spectrum, antenna-pattern shaping, or multiple scanning-beam concept can do better. Inner bounds are found by construction. The bounds provide a yardstick against which the performance of any system can be compared. Detailed results of the effectiveness of frequency reuse are presented for the important case of two closely spaced ground stations.*

I. INTRODUCTION

Communication satellites employing large antennas capable of providing multiple, nonoverlapping, high-gain spot beams have attracted considerable attention in recent years because the allocated spectral bands can be reused in the various spot beams.[1–17] For a fixed level of available RF power, such satellites exhibit a much higher throughput capability relative to area coverage satellites because the higher antenna gain reduces the RF power required per spot beam to maintain the same effective radiated power (and therefore the same communication performance) on the earth; the excess power can then be applied to the remaining spot beams, thereby increasing the capacity of the satellite. For widely separated spot beams, each such beam can be considered as providing an independent, full-bandwidth channel between the satellite and the earth. A satellite offering N independent spot beams thereby provides a factor of N higher capacity than does an area coverage satellite. In fact, the total RF power required by the multiple spot beam satellite to provide this increased capacity is often less than that needed by the lower capacity area coverage system because the number of spot beams provided is usually smaller than the ratio of the spot beam to area coverage antenna gain.

In this paper, we first briefly review several spot-beam, frequency-reuse concepts proposed to date. These include the familiar satellite-switched, time-division multiple access (SS-TDMA) approach with fixed spot beams,[1–9] integration of the basic SS-TDMA structure with a rapidly scannable spot beam to provide service over those regions not covered by a fixed spot beam,[10–15] and multiple scanning beam approaches which efficiently accommodate nonuniform service demand across some wide service region.[14,17] This background material appears in Section II and is optional reading for those already familiar with these concepts.

A fundamental question unanswered by these approaches concerns the ultimate capacity limit of frequency reuse communication satellite networks characterized by multiple ground stations arbitrarily located over some wide service area and by constraints upon the space platform radiated power, the available bandwidth, and the physical dimensions of the satellite's antenna aperture. This question is addressed from an information theoretic point of view in Section III, and inner and outer bounds on the capacity limit are derived. Because the satellite integrates multiple users with nonuniform traffic demands, the capacity limit is expressed as a region over the space of allowable transmission rates $\mathscr{R}_1, \cdots, \mathscr{R}_G$ such that independent information can be reliably transmitted at rate $\mathscr{R}_1$ to ground station number 1, rate $\mathscr{R}_2$ to ground station number 2, $\cdots$, and rate $\mathscr{R}_G$ to ground station number G. The G-dimensional region of acceptable rates $\mathscr{R}_1, \cdots, \mathscr{R}_G$ defines the capacity region of the multiuser satellite system. For any traffic demand within this region, the power, bandwidth, and antenna aperture resources of the satellite can always be apportioned to satisfy the demand. The derived bounds thereby provide a yardstick against which the capabilities of any system concept can be compared and also suggest areas in systems, coding, antennas, and devices where additional research might greatly enhance the capabilities of future satellite systems. It is noted that the outer bound is all-inclusive; that is, no TDMA, FDMA, spread-spectrum, antenna-pattern shaping, or scanning spot-beam concept can provide a capacity region exceeding this bound. Attention is restricted to the down-link, although similar bounds can be derived for the up-link as well. We note that the Fourier transform relationship between the antenna aperture distribution and the far-field radiation pattern implies that the geographical distances between ground stations affects the capacity region.

To find an outer bound on the capacity region, cooperation among the information sources is permitted at the satellite, and cooperation among ground stations is also permitted. Subject to radiated power, bandwidth, and satellite antenna aperture constraints, the total capacity is found for all combinations of G, G-1, $\cdots$, 2, 1 ground stations in the network. For example, if there are two ground stations, the capacities into each individually are found, along with the combination of both treated as a single receiver. Call these C_1, C_2, and $C_{1,2}$. Then the capacity region outer bound is the intersection of the regions $\mathscr{R}_1 < C_1$, $\mathscr{R}_2 < C_2$, $\mathscr{R}_1 + \mathscr{R}_2 < C_{1,2}$.

By definition, any realizable approach for which the capacity region can be found provides an achievable inner bound to the capacity region. Transmission via independent beams in the absence of both cooperation among message sources at the satellite and cooperation among receivers provides one such bound. Each link from the satellite to a ground station is corrupted by Gaussian noise and possibly cochannel interference from neighboring beams. By varying the power division among beams, an inner bound is swept out. Time-sharing of several movable beams provides another inner bound. Tight inner bounds can often be obtained through judicious selection of both the number of beams chosen and their associated aperture distributions. It is noted that construction of the inner bound also provides a method for achieving the inner bound.

Thus, in Section III we extend the concept of system capacity to multiuser satellite systems by extending the usual power and bandwidth constraints for a single-user channel to include an aperture-size constraint as well. The approach differs from earlier multiple-user analysis[18–21] in that here we are free to select the number of beams, their radiation patterns, and the apportionment of power among beams to maximize capacity. We note that these bounding techniques can readily be applied to a more restrictive satellite antenna wherein arbitrary aperture distributions subject only to aperture extent limitations are replaced by other restraints. (For example, one might require that the aperture contain an N-element, phased-array antenna of fixed element spacing and fixed elemental radiation pattern.)

In Section IV, we invoke the theory of discrete prolate spheroidal sequences to numerically evaluate capacity region outer bounds for systems containing two closely spaced ground stations. This important case is examined in detail because, for two closely spaced ground stations, it has generally been assumed that time-sharing a single beam, rather than frequency reuse, must be employed to serve both. We can now, for the first time, investigate this hypothesis. Although analytical results can readily be obtained for systems containing three or more ground stations, such results cannot easily be displayed graphically (three or more dimensions are needed to portray the capacity region), and provide little physical insight into the utility of frequency reuse. Consequently, attention is restricted to networks containing two closely spaced ground stations. A typical result is as follows. Suppose the two ground stations are located at the −1.3-dB contours (one on each side of center) of the narrowest beam which can be produced by the satellite. Suppose further that the maximum available carrier-to-noise ratio is 15 dB; that is, we could obtain a 15-dB carrier-to-noise ratio at one ground station if all available power is given to one beam with maximum gain in the direction of that ground station. Finally, both ground stations require the same transmission rate $\mathscr{R} = 3.5$ b/s/Hz. Then two simple pattern-shaped beams are found, one serving each of the two ground stations and each reusing the total available bandwidth, such that the power required is within 1.5 dB of the outer bound for that particular ground-station separation. Thus, the most complicated coding and radiation scheme possible could save, at most, 1.5 dB in required total power to achieve the rate $\mathscr{R}$ provided by these two beams. Moreover, time-sharing a single scannable beam between the two ground stations would require 6 dB more total power to achieve the same capacity.

In Section V, we qualitatively explore additional implications of the satellite system capacity region. Addressed here are the effects of rain fading, resource sharing to combat rain fading, and point-to-multipoint or broadcast satellite considerations.

II. BACKGROUND

The earliest attempts at employing multiple spot-beam, frequency-reuse concepts involved the formation of a number of up-link and down-link spot beams permanently directed to a corresponding number of geographically isolated service regions which might appear as shown in Fig. 1a.[1-9] To interconnect the various spot-beam regions, an on-board time-division satellite switch is employed as shown in Fig.

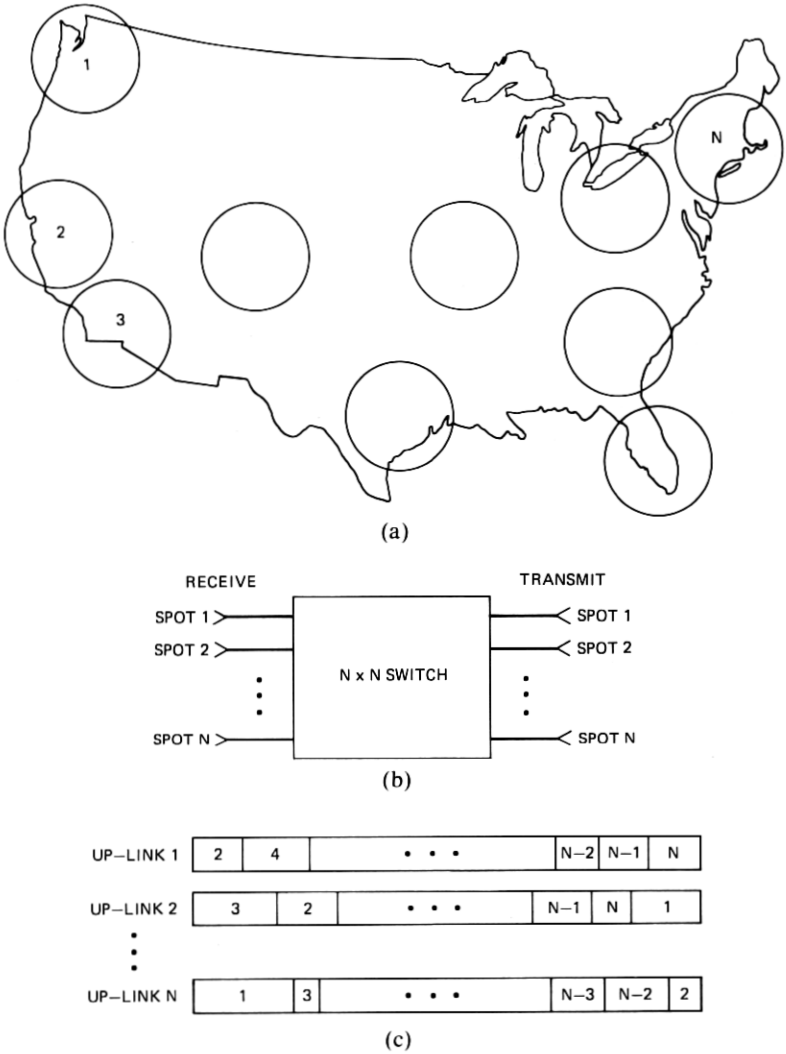

Fig. 1—Fixed-beam satellite switched time-division multiple access. (a) Typical fixed-beam service regions. (b) Time-division satellite interconnecting switch. (c) Typical interconnection sequences.

1b; a time sequence of the various interconnections made by the satellite switch might then appear as shown in Fig. 1c. Within each interconnection time slot, the ground stations of the network communicate via time-division multiple access (TDMA). Digital modulation consistent with TDMA is employed. In this manner, the capacity of the satellite is time-shared among users.

To maintain cochannel interference among beams at an acceptably low level, the fixed spot-beam service regions must be spatially separated by perhaps two or more 3-dB beamwidths of the spot beam radiation pattern; closer spacing requires that a greater level of RF power be made available for each spot beam to overcome the degradation resulting from interbeam interference. Hence, the service regions are often separated by a much larger area within which no user can access the satellite (blackout region).

Several approaches intended to reduce or eliminate the blackout region have been proposed. One technique consists of splitting the spectral band into several subbands and assigning these to spot-beam regions such that reuses of the same subband are separated by a distance sufficient to satisfy the interference criteria.[16] Additional spot beams are then provided using different subbands to cover the blackout region. Unfortunately, the demand for service can exhibit wide variations among the spot-beam regions, and the band-splitting approach limits the available capacity into any one spot-beam service region to that provided by one subband.

Another approach employs a pair of steerable spot beams (one for the up- and one for the down-link) in conjunction with fixed spot beams to provide service to the blackout region.[10] At each point in time, the up-link of the scanning beam can receive messages from a small spot-beam region contained within the blackout region, and the down-link can deliver messages to another small region within the blackout region. The scanning beam antennas appear as additional ports on the satellite switch.[15,17] Then, by scanning the up- and down-links in such a manner that the amount of time dwelled at any one location is proportional to the traffic for that region, the traffic of the entire blackout region is accommodated by the single scanning spot-beam antenna. Thus, the scanning-beam concept combines the wide-access capability of an area coverage system with the power saving afforded by a spot-beam system.

A drawback of the scanning plus multiple spot-beam concept arises from the fact that the demand for service from all regions (with the former blackout region now being treated as one large region) are dissimilar. Each region can now be served by the total available spectrum, but not all regions may require this capacity, implying waste of satellite resources. The total RF power available at the satellite can, however, be apportioned among the service regions on the basis of capacity requirements, thereby improving utilization efficiency somewhat.

Yet another approach involves the use of multiple scanning beams.[14,17] Here, each antenna port is capable of providing a narrow spot beam that can be scanned anywhere over the total service region.

With this approach, it is possible to tailor the dwell time of a spot beam at any spot-beam region in proportion to the demand for service from that region. This concept is quite efficient in that a limited number of transponders are shared among a far larger number of spot-beam regions, thereby affording the power saving associated with highly directive spot-beam antennas while providing service over a wide region characterized by nonuniform traffic. The required number of spot beams is readily established by requiring that the total offered traffic (measured in time-slot units) equal the total number of available time slots. However, by employing a larger number of beams and transponders, each using the full bandwidth, one can envision the use of channel coding techniques within each beam to save power by using the available excess bandwidth. The number of scanning beams which can be provided before interbeam interference becomes a problem is still an open question, however. Because the dwell time at any one location is dependent upon the traffic from that location, this number must be dependent upon the terrestrial traffic distribution.

Thus we see that various approaches have been proposed to introduce spot-beam frequency-reuse techniques for satellite communications. These represent varying degrees of refinement of the basic approach in an attempt to more efficiently allocate the power and bandwidth resources of the satellite such that a uniform grade of service is maintained over a service area much larger than the beamwidth of a single spot beam. The traffic patterns might be highly nonuniform over the total service area.

III. DERIVATION OF THE CAPACITY REGION BOUNDS

The satellite system to be studied is shown in Fig. 2. Here, the satellite is presented with G independent messages M_1 through M_G to be communicated, respectively, to geographically separated identical

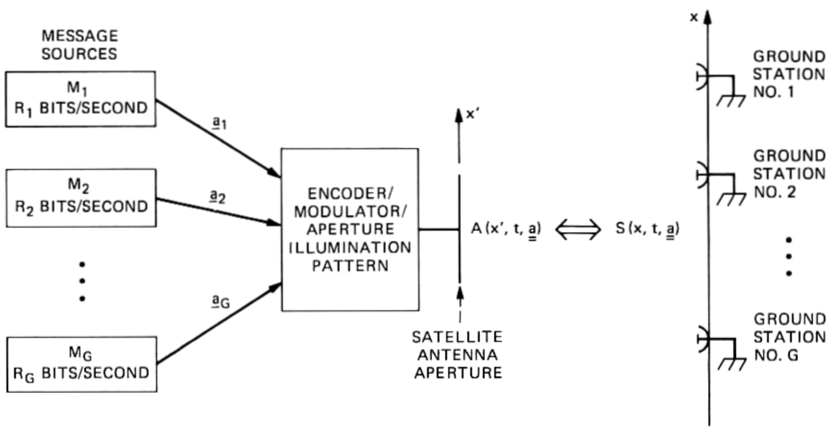

Fig. 2—Satellite system model.

ground stations 1 through G. The receiving gain of each ground station is normalized to unity. These messages are typically produced at the satellite via onboard detection of the up-links. Thus, message source M_j, $1 \leq j \leq G$, represents the composite of all up-link communication from the various ground stations intended for reception at ground station number j. It is assumed that the various up-link messages are detected at the satellite error-free. This is consistent with up-link transmission at rates lower than those permitted by the up-link capacity region. Since power constraints are usually not as tight on the up-link as on the down-link, we expect the satellite system capacity to be dominated by the down-link capacity region, and bounds are found only for the down-link; similar bounds can be established for the up-link, however. Message source M_j produces information at a rate of $\mathscr{R}_j$ b/s, $j = 1, \cdots, G$. The satellite is constrained to a total radiated power of P watts and an RF radiation bandwidth of W Hz. The antenna aperture distribution is identically zero outside some two-dimensional region in the plane of the aperture.

3.1 Outer bound

We seek first an outer bound on the collection of rate vectors $\underline{\mathscr{R}} = [\mathscr{R}_1, \cdots, \mathscr{R}_G]$ for which reliable communication from the satellite to the ground stations is possible. Within this region, the power, bandwidth, and aperture dimension resources of the satellite can be allocated to achieve error-free communication at rates $\underline{\mathscr{R}}$.

Let message source no. i produce the binary information stream $\underline{a}_i$, $i = 1, \cdots, G$. The collection of information streams can then be represented by the matrix $\underline{\underline{a}} = [\underline{a}_1, \cdots, \underline{a}_G]$. For a particular message matrix $\underline{\underline{a}}$, the aperture distribution at the satellite can be represented as

$$\mathscr{A}(x', t, \underline{\underline{a}}) = \mathrm{Re}[\tilde{\mathscr{A}}(x', t, \underline{\underline{a}})e^{j\omega_0 t}], \tag{1}$$

where ω_0 is the carrier frequency and $\tilde{\mathscr{A}}$ is the complex baseband aperture distribution. In the above, we have, for simplicity, assumed a one-dimensional antenna; the approach can be generalized to a planar antenna, however. The far-field radiation pattern $S(x, t, \underline{\underline{a}})$ is obtained via a Fourier transform of the aperture distribution. For baseband bandwidth small compared to the carrier frequency ω_0,

$$S(x, t, \underline{\underline{a}}) = \mathrm{Re}\left[\frac{1}{2\pi}\int \tilde{\mathscr{A}}(x', t, \underline{\underline{a}})e^{jkxx'/z_0}e^{j\omega_0 t}\, dx'\right], \tag{2}$$

where $k = \omega_0/c$, c is the speed of light, and z_0 is the distance between the plane of the satellite and the plane of the earth. Equation (2) applies in the far field, i.e., $z_0 \gg$ aperture extent.

Let

$$\bar{A}(x\,t,\,\underline{a}) \Leftrightarrow \bar{\mathscr{A}}(x',\,t,\,\underline{a}), \tag{3}$$

where ($\Leftrightarrow$) implies the Fourier transform. Then the farfield radiation pattern is:

$$S(x,\,t,\,\underline{a}) = A_r(x,\,t,\,\underline{a})\cos\omega_0 t + A_i(x,\,t,\,\underline{a})\sin\omega_0 t, \tag{4}$$

where A_r and A_i are, respectively, the real and imaginary components of $\bar{A}$.

We see that the far-field radiation contains in-phase and quadrature components. Again, because the baseband bandwidth is small compared to the carrier frequency, these two components are orthogonal. Thus, the in-phase and quadrature components can provide two independent, identical capacity channels. Attention is therefore restricted to one channel, say, the in-phase channel; results for the quadrature channel are identical.

The ground stations are located at $x_1, \cdots, x_G$ (again, for simplicity, we assume that ground stations are colinear; the approach can be extended to planar constellations, however). The ground stations are allowed to communicate via a noiseless terrestrial link to a central processor as shown in Fig. 3. At the ith ground station, we receive

$$\mathscr{Y}(x_i,\,t,\,\underline{a}) = A_r(x_i,\,t,\,\underline{a})\cos\omega_0 t + n_i(t)\cos\omega_0 t, \tag{5}$$

where $n_i(t)$ is white, normal, and of spectral density $N_0/2$. We see,

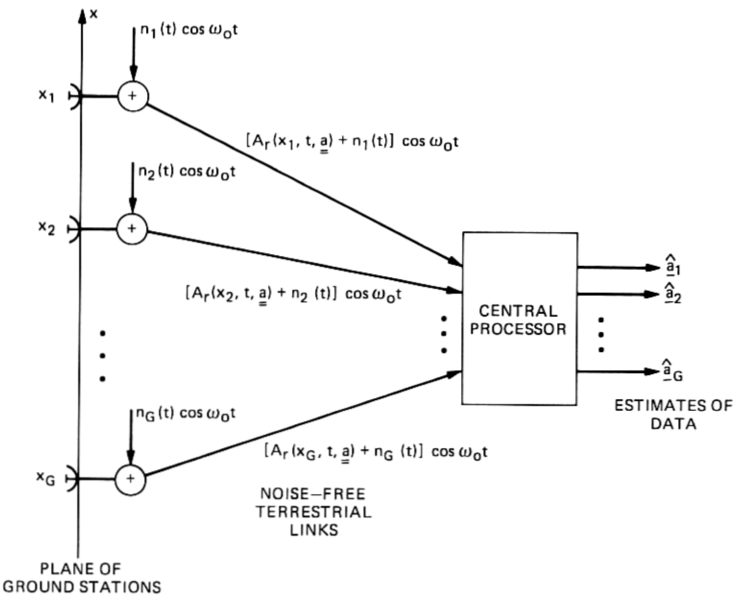

Fig. 3—Processing model for capacity region outer bound.

then, that the central processor is presented with the G-dimensional received vector

$$\underline{\mathcal{Y}} = \left\{ \begin{bmatrix} A_r(x_1, t, \underline{\underline{a}}) + n_1(t) \\ A_r(x_2, t, \underline{\underline{a}}) + n_2(t) \\ \vdots \qquad\qquad \vdots \\ A_r(x_G, t, \underline{\underline{a}}) + n_G(t) \end{bmatrix} \right\} \cos \omega_0 t, \tag{6}$$

where n_i, n_j are independent, $i \neq j$.

Let us express $A_r(x_i, t, \underline{\underline{a}})$ as an expansion in some complete orthonormal basis $\{u(t)\}$:

$$A_r(x_i, t, \underline{\underline{a}}) = \sum_j b_j(x_i, \underline{\underline{a}}) u_j(t). \tag{7}$$

Then, the part of the received vector relevant to detection is

$$\underline{Z} = \left\{ \sum_j \begin{bmatrix} b_j(x_1, \underline{\underline{a}}) \\ \vdots \\ b_j(x_G, \underline{\underline{a}}) \end{bmatrix} u_j(t) + \sum_j \begin{bmatrix} n_{1,j} \\ \vdots \\ n_{G,j} \end{bmatrix} u_j(t) \right\} \cos \omega_0 t, \tag{8}$$

where

$$n_{i,j} = \int n_i(t) u_j(t)\, dt, \qquad i = 1, \cdots, G. \tag{9}$$

Clearly,

$$E\{n_{i,j} n_{k,p}\} = \frac{N_0}{2} \delta_{i,k} \delta_{j,p}. \tag{10}$$

Let vectors $\underline{\ell}_1, \cdots, \underline{\ell}_G$ be an orthonormal basis spanning G-dimensional Euclidean vector space. Then, for a particular set of ground station locations $x_1, \cdots, x_G$, and for every message matrix $\underline{\underline{a}}$, we can expand $\underline{Z}$ in terms of the orthonormal basis $\{\underline{\ell}\}$:

$$\underline{Z} = \left\{ \sum_j \sum_{i=1}^{G} K_{i,j}(\underline{\underline{a}}) u_j(t) \underline{\ell}_i + \sum_j \sum_i \eta_{i,j} u_j(t) \underline{\ell}_i \right\} \cos \omega_0 t, \tag{11}$$

where

$$\eta_{i,j} = \begin{bmatrix} n_{1,j} \\ \vdots \\ n_{G,j} \end{bmatrix} \cdot \underline{\ell}_i \tag{12a}$$

and

$$E(\eta_{i,j} \eta_{k,p}) = \frac{N_0}{2} \delta_{i,k} \delta_{j,p}. \tag{12b}$$

In (11), the expansion coefficients $K_{i,j}(\underline{\underline{a}})$ are dependent upon the particular basis $\{\underline{\ell}\}$ employed.

Thus, we see that since reception of the radiated signal can occur at no more than G points, the total channel from the satellite to the ground stations consists of G independent parallel channels, each of bandwidth W. Viewed from the perspective of the G ground terminal receivers, the satellite is, in effect, radiating G-dimensional vectors rather than continuous functions of space since reception can only occur at the G points where receivers are located. We conclude that the most general baseband radiation to maximize the capacity of the G parallel satellite channels is of the form

$$\mathcal{s}(t, x, \underline{a}) = \sum_j \sum_{i=1}^{G} \xi_{i,j}(\underline{a})\psi_i(x)u_j(t), \tag{13}$$

and the signal portion of the received vector is of the form

$$\underline{S}(t, \underline{a}) = \sum_j \sum_{i=1}^{G} \xi_{i,j} \begin{bmatrix} \psi_i(x_1) \\ \vdots \\ \psi_i(x_G) \end{bmatrix} u_j(t). \tag{14}$$

In (14), the vectors $\underline{\psi}_i = \begin{bmatrix} \psi_i(x_1) \\ \vdots \\ \psi_i(x_G) \end{bmatrix}$ and $\underline{\psi}_j = \begin{bmatrix} \psi_j(x_1) \\ \vdots \\ \psi_j(x_G) \end{bmatrix}$ are orthogonal, $i \neq j$. For convenience, we normalize the energy of the underlying radiation patterns to unity, i.e., $\int |\psi_i(x)|^2 \, dx = 1$, $i = 1, \cdots, G$. The functions $\psi_i(x)$ and $\psi_j(x)$ are not necessarily orthogonal, however, over the real line, $i \neq j$.

We note from (14) that some radiation patterns will deliver more radiated power to the G ground stations than others. The problem, then, is to find G radiation patterns $\psi_i(x)$, producing G orthogonal receiving vectors $\underline{\psi}_i$, which maximize the total capacity, subject to the constraint imposed by the finite radiation aperture of the satellite antenna. This maximization is performed under the additional constraint that the total radiated power is bounded. Since the G receiving vectors are orthogonal, selecting their inputs to be independent maximizes the capacity, i.e.,

$$E\{\xi_{i,j}\xi_{k,p}\} = E\{\xi_{i,j}^2\}\delta_{i,k}. \tag{15}$$

Let the total power radiated by the satellite be constrained to P watts. Thus, for each $\underline{a}$,

$$\frac{1}{T}\iint \mathcal{s}^2(t, x, \underline{a}) \, dt \, dx \leq P, \tag{16}$$

where T is the temporal extent of the transmission. Since we are interested in obtaining an outer bound to the capacity region, any bound derived from a constraint which contains (16) as a special case must be an outer bound to the true capacity region. Rather than insisting that (16) hold for each $\underline{a}$, let us impose the constraint that, averaged over all information matrices, the radiated power must not exceed P, i.e.,

$$\frac{1}{T}\int\int \langle s^2(t, x, \underline{a})\, dt\, dx\rangle \qquad \le P. \tag{17}$$

Substituting (13) and invoking the orthonormality of the set $\{u(t)\}$, the normality of the set $\{\psi(x)\}$, and the independence of the encoded symbols $\{\xi\}$, we obtain:

$$\sum_j \sum_{i=1}^{G} \langle \xi_{i,j}^2 \rangle = P. \tag{18}$$

For each i, we can interpret $\sum_j < \xi_{i,j}^2 >$ to be the power in the ith mode. For convenience, let

$$\sum_j \langle \xi_{i,j}^2 \rangle = \alpha_i P \tag{19}$$

$$\sum_{i=1}^{G} \alpha_i = 1. \tag{20}$$

Each mode can radiate both in-phrase and quadrature components of the received process. Thus, the ith mode provides an additive white Gaussian noise (AWGN) channel of bandwidth W and power $\alpha_i P$. Invoking the classical formula for the AWGN channel, the total capacity for the G parallel channels is seen to be

$$C_T^G = \sum_{i=1}^{G} W \log_2 \left\{ 1 + \frac{\alpha_i P \sum_{k=1}^{G} \psi_{i,k}^2}{N_0 W} \right\}, \tag{21}$$

where $\psi_{i,k}$ is the kth component of the vector

$$\underline{\psi}_i = \begin{bmatrix} \psi_i(x_1) \\ \vdots \\ \psi_i(x_G) \end{bmatrix} \tag{22}$$

and $\psi_i(x)$ is the normalized radiation pattern corresponding to the ith mode. In writing (21), we have assumed ground-station receiving apertures normalized to unity. This normalization is permissible since we will soon rewrite (21) in terms of the available carrier-to-noise ratio (CNR) at a receiving terminal. We note that, for all i, k, the vectors $\underline{\psi}_i$ and $\underline{\psi}_k$ are orthogonal, i.e.,

$$\underline{\psi}_i \cdot \underline{\psi}_k = |\underline{\psi}_i|^2 \delta_{i,k}, \tag{23}$$

but that the functions $\psi_i(x)$ and $\psi_k(x)$ need not be orthogonal.

For each G-member set of functions $\{\psi(x)\}$ and for each power allocation $\alpha_1, \cdots, \alpha_G$, the total capacity is given by (21). We note that this can be maximized by allocating power according to the classical parallel channel water-fill problem.[22] Here, we consider that G Gaussian parallel channels are available, of noise powers

$$\sigma_i^2 = \frac{N_0 W}{\sum_{k=1}^{G} \psi_{i,k}^2}, \qquad i = 1, \cdots, G. \tag{24}$$

The total available power is P, and we must select $\alpha_1, \cdots, \alpha_G$ satisfying the constraint (20) such that the capacity is maximized. The solution is

$$\sigma_i^2 + \alpha_i P = B \quad \text{for } \sigma_i^2 < B \tag{25}$$

$$\alpha_i = 0 \quad \text{for } \sigma_i^2 > B, \tag{26}$$

and B is chosen such that (20) is satisfied. This solution has the geometrical interpretation shown in Fig. 4. Available is a quantity of water P to be placed in a vessel with a bottom consisting of J levels as shown ($J = 6$ in Fig. 4). The height of the ith level is the noise power σ_i^2. Then, the amount of power given to the ith channel is equal to the height of water above the ith level. For any channels exhibiting too high a noise power for the available signal power, the greatest capacity is achieved by allocating no power to those channels.

The G normalized functions $\psi_1(x), \cdots, \psi_G(x)$ were chosen such that (23) is satisfied for all i, k and are also aperture-limited, that is, the Fourier transforms of these functions vanish outside the finite dimensions of the antenna aperture but are otherwise arbitrary. To maximize

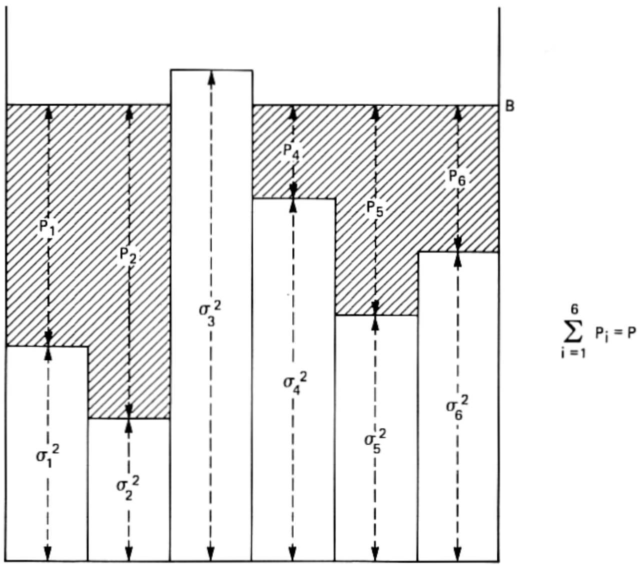

Fig. 4—Parallel channel water-fill analogy.

the capacity of the G parallel channels, we must select those modes or functions $\psi_1(x), \cdots, \psi_G(x)$ satisfying the constraints and must further allocate power among the modes such that (21) is maximized. It is convenient to first place (21) in a form involving available CNR at a receiving terminal.

Consider an antenna aperture limited in extent to $[-A, A]$. The maximum gain available from this antenna in the direction of a ground station located at $x = 0$ arises for the normalized aperture function:

$$\Phi(x') = \begin{cases} \sqrt{\pi/A}, & -A < x' < A \\ 0, & \text{elsewhere.} \end{cases} \tag{27}$$

For this function,

$$\psi(x) = \sqrt{\frac{A}{\pi}} \frac{\sin kAx/z_0}{kAx/z_0}. \tag{28}$$

Thus, if all power is given to this mode, the power available at a ground station located at $x = 0$ is $Pg_{\max}$, where

$$g_{\max} = \frac{A}{\pi}. \tag{29}$$

For this case, the carrier-to-noise ratio is simply

$$\rho = \frac{Pg_{\max}}{N_0 W}. \tag{30}$$

Using the relationship (30), the total capacity for the G ground terminal network is given by

$$C_T^G = \underset{\{\alpha\},\{\psi(x)\}}{\text{supremum}} \left\{ \sum_{i=1}^{G} W \log_2 \left[1 + \rho\alpha_i \frac{\sum_{k=1}^{G} \psi_{i,k}^2}{g_{\max}} \right] \right\}, \tag{32}$$

where

$$\int \psi_i^2(kx/z_0)\, d(kx/z_0) = 1, \qquad i = 1, \cdots, G \tag{33}$$

$$\int_{\infty}^{\infty} \psi_i(kx/z_0) e^{-jxx'/z_0} d(kx/z_0) = 0, \qquad |x'| > A \tag{34}$$

$$\underline{\psi}_i = \begin{bmatrix} \psi_i(x_1) \\ \vdots \\ \psi_i(x_G) \end{bmatrix} \tag{35}$$

$$\alpha_i \alpha_\ell \underline{\psi}_i \cdot \underline{\psi}_\ell = \alpha_i^2 |\underline{\psi}_i|^2 \delta_{i,\ell} \tag{36}$$

$$\sum_{i=1}^{G} \alpha_i = 1 \tag{37}$$

$$g_{\max} = \frac{A}{\pi}. \tag{38}$$

Clearly, the location of the ground stations $x_1, \cdots, x_G$ and the antenna aperture $2A$ affect the maximum capacity. We note from (36) that α_i and α_ℓ are included in the orthogonality constraint because the radiation vector of any mode receiving zero power (i.e., $\alpha = 0$ for that mode) need not satisfy any orthogonality conditions.

Condition (34) implies that, to maximize (32), each $\psi_i(x)$ can be expressed as some linear combination of G basis functions, each itself a linear combination of the G sampling functions $[\sin kA(x - x_\ell)/z_0]/kA(x - x_\ell)/z_0]$ $\ell = 1, \cdots, G$, with the property that the kth basis function equals unity at $x = x_k$ and equals zero at x_ℓ, $\ell \neq k$.[23] The coefficients in the expansion of each $\psi_i(x)$ are then chosen, subject to (36), such that (32) is maximized. When the ground stations are uniformly spaced, the theory of discrete prolate spheroidal sequences[24,25] can be applied to more directly yield the capacity.

Having found the total capacity possible for a G-terminal network with power, bandwidth, and aperture dimension constraints and with noiseless channels connecting each ground station to a central processor unit, we conclude that the sum of the information rates into the G-terminals cannot exceed C_T^G, i.e.,

$$\mathscr{R}_1 + \mathscr{R}_2 + \cdots + \mathscr{R}_G \leq C_T^G. \tag{39}$$

The capacity region must therefore be totally contained within the region (39).

Proceeding in an identical manner, we can find the maximum capacity possible into a network containing any G-1 of the original ground stations, assuming that all of the available satellite resources are available for communicating to the G-1 terminals chosen. In this manner, we find G capacities $C_{T,m}^{G-1}$, $m = 1, \cdots, G$, one capacity for each choice of G-1 out of G participating ground stations. We conclude that the capacity region must be totally contained within the region.

$$\mathscr{R}_1 + \cdots + \mathscr{R}_{G-1} < C_{T,1}^{G-1} \tag{40}$$

$$\mathscr{R}_1 + \cdots + \mathscr{R}_{G-2} + \mathscr{R}_G \leq C_{T,2}^{G-1} \tag{41}$$

$$\vdots$$

$$\mathscr{R}_1 + \mathscr{R}_3 \cdots + \mathscr{R}_G \leq C_{T,G-1}^{G-1} \tag{42}$$

$$\mathscr{R}_2 + \mathscr{R}_3 + \cdots + \mathscr{R}_G \leq C_{T,G}^{G-1}. \tag{43}$$

Continuing, we find the maximum capacity for all combinations of G-2, G-3, $\cdots$, 2, 1 ground stations in the network, and write an inequality for the sum of the rates for each combination. An outer bound on the capacity region for the G-terminal network is then the intersection of the regions corresponding to each inequality. For example, if $G = 2$, we can find three regions,

$$\mathscr{R}_1 + \mathscr{R}_2 \leq C_T^2 \tag{44}$$

$$\mathscr{R}_1 \leq C_{T,1}^1 \tag{45}$$

$$\mathscr{R}_2 \leq C_{T,2}^1. \tag{46}$$

The intersection of these three regions, shown in Fig. 5, is an outer bound on the capacity region of the two-terminal network.

We note that the outer bound (32) can be applied to other aperture constraints more restrictive than the size constraint (34). For example, one might require that each member of the set $\{\psi(x)\}$ be realized by a phased-array antenna with a specified-element radiation pattern and a fixed-element separation. Then each member of the set $\{\psi(x)\}$ would be represented as a linear combination of the radiation patterns of the elements, and C_T^G would be represented as the supremum over the coefficients in each such linear expansion and the set $\{\alpha\}$. The coefficients must then satisfy a normalization relationship such that (33) is satisfied. Alternatively, the antenna might be specified as being of offset Cassegrainian design with fixed-feed arrangement, and each member of the set $\{\psi(x)\}$ would be represented as a linear combination of the radiation from each feed. Other constraints are, of course, also possible. For any other set of constraints, however, the resulting capacity region would be smaller than that obtained via (32) through (38), i.e., the capacity region obtained via (32) through (38) contains, as a subregion, the capacity region obtained by a constraint more restrictive than a size constraint.

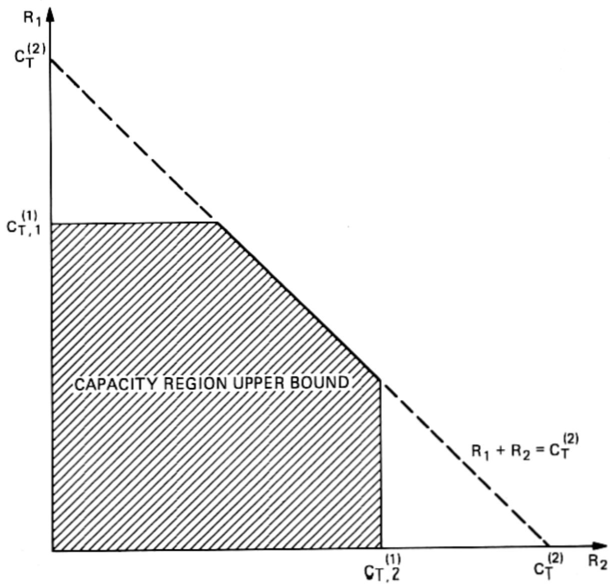

Fig. 5—Typical outer bound on the capacity region for two ground stations.

3.2 Inner bound

We now turn our attention to obtaining a lower bound on the capacity region of the G-terminal network. To find a lower bound, we remove the central processing unit of Fig. 3 as well as the noiseless terrestrial interconnecting links. Any method for then communicating the informational contents of message sources M_1 through M_G to the intended destinations provides an inner bound to the capacity region if the capacity region for that method can be calculated. The tightness of the bound varies with the particular method chosen.

One method for communicating between the satellite and the ground stations consists of time-sharing a single channel and antenna beam among all destinations. The antenna can form a single beam, directed anywhere, of maximum gain $g_{\max}$; when focused at one particular receiving terminal, the available carrier-to-noise ratio is ρ. The single channel capacity for this transmission scheme is then

$$C_1 = W \log[1 + \rho]. \tag{47}$$

By varying the amount of time that this single beam dwells at each ground station in proportion to the traffic demand for that station, we obtain the following lower bound for the capacity region:

$$\mathcal{R}_1 + \cdots + \mathcal{R}_G \leq C_1 = W \log[1 + \rho]. \tag{48}$$

Another inner bound can be found by having the satellite form G beams, each of maximum gain in the direction of one ground station. The ith beam is given $\alpha_i P$ watts of power, $\sum_{i=1}^{G} \alpha_i = 1$. To transmit at capacity, the encoded input symbols to each beam are independently drawn from a Gaussian distribution of variance $\alpha_i P$, $i = 1, \cdots, G$. The spectrum is totally reused in the various beams, and the channel inputs to all beams are independent.

The pth ground station observes both the desired signal as well as independent additive Gaussian interference from all other beams; the strengths of these interferers are dependent upon the locations of the ground stations. Let $\kappa_{p,i}$ be the power from the ith interferer observed at the pth ground station relative to the desired signal at the pth ground station. For maximum gain beams,

$$\kappa_{p,i} = \left[\frac{\sin\{kA(x_p - x_i)/z_o\}}{\{kA(x_p - x_i)/z_o\}} \right]^2, \qquad i = 1, \cdots, G; \qquad p = 1, \cdots, G. \tag{49}$$

The rate at which information can be reliably communicated into the pth ground station is then bounded by

$$\mathcal{R}_p \leq W \log\left[1 + \frac{\alpha_p \rho}{1 + \rho \sum_{\substack{i=1 \\ i \neq p}}^{G} \alpha_i \kappa_{p,i}} \right], \qquad p = 1, \cdots, G. \tag{50}$$

For a fixed allocation of power to the various beams, the point $\underline{\mathscr{R}}$ obtained via (50) with equality provides an achievable rate vector and is therefore a point on an inner bound to the capacity region. By varying the power allocation among beams, an inner bound on the capacity region is swept out.

Yet another inner bound is obtained by again forming G independent beams, each of maximum gain in the direction of one ground station subject to the constraint that no interference be radiated in the direction of any other ground station. The encoded channel inputs to each beam are selected from independent Gaussian distributions. Thus, G independent, noninterfering channels are created. Again, the noninterfering constraint implies that the maximum gain for each beam formed is dependent upon ground-station locations. Non-interfering beams are found as follows.

Consider the pth ground station. Let $Y_p(x')$ be the aperture distribution for the pth beam. Then the radiated field for this distribution is

$$u_p(x) = \frac{1}{2\pi} \int_{-1}^{1} Y_p(x') e^{jxx'}\, dx', \tag{51}$$

where, for convenience, we have normalized the aperture dimension A and the scale factor k/z_o both to unity. Now we require that

$$u_p(x_j) = 0, \qquad j \neq p. \tag{52}$$

Using Lagrange multipliers $a_j, j \neq p$, we wish to maximize

$$|\, u_p(x_p)\,|^2 = \left| \frac{1}{2\pi} \int_{-1}^{1} \psi_p(x') \left[e^{jx_p x'} - \sum_{\substack{j=1 \\ j\neq p}}^{G} a_j e^{jx_j x'} \right] dx' \right|^2. \tag{53}$$

Invoking the Schwartz inequality, we find the maximizing aperture distribution to be

$$\psi_p(x') = B \left[e^{-jx_p x'} - \sum_{\substack{j=1 \\ j\neq p}}^{G} a_j e^{-jx_j x'} \right], \tag{54}$$

where the constants $[a_j,\ j \neq p\}$ must be selected to satisfy the constraints (52) and B is selected such that $\int |\, u_p(x)\,|^2\, dx = 1$. From the constraint equations (52),

$$u_p(x_j) = 0 = \frac{B}{2\pi} \int_{-1}^{1} \left[e^{-jx_p x'} - \sum_{\substack{k=1 \\ k\neq p}}^{G} a_k e^{-jx_k x'} \right] e^{jx_j x'}\, dx \tag{55}$$

$$\Rightarrow \frac{\sin(x_j - x_p)}{x_j - x_p} = \sum_{\substack{k=1 \\ k\neq p}}^{G} a_k \frac{\sin(x_j - x_k)}{x_j - x_k}, \qquad \begin{matrix} j = 1, \cdots, G. \\ j \neq p \end{matrix} \tag{56}$$

The G-1 equations (56) can be solved for the unknown parameters a_j, $j \neq p$, and the normalization factor B follows. The remaining G-1 noninterfering beams are found similarly. For a given power allocation $\{\alpha\}$, we then obtain

$$\mathscr{R}_j \leq W \log\left[1 + \alpha_j \rho \frac{g_j(x_j)}{g_{\max}}\right], \qquad j = 1, \cdots, G, \tag{57}$$

where $g_j(x_j) \leq g_{\max}$ is the gain of the jth mode at the location of the jth ground station. Again, by varying the power allocation, we sweep out a lower bound on the capacity region.

The three inner bounds for $G = 2$ might appear as shown in Fig. 6. We note that, depending on the ground station locations, the multibeam bounds may or may not exceed the time-share bound. Other inner bounds can also be found. Of course, none of these bounds can exceed the previously found outer bound.

We note that, by construction, derivation of an inner bound yields a method whereby the inner bound can be attained since, when deriving these bounds, we did not require cooperating ground terminals.

IV. DETAILED STUDY OF TWO GROUND TERMINAL SYSTEMS

The bounding techniques introduced in the previous section are now applied to study the capacity region for two ground terminal systems subject to power, bandwidth, and antenna aperture dimensional con-

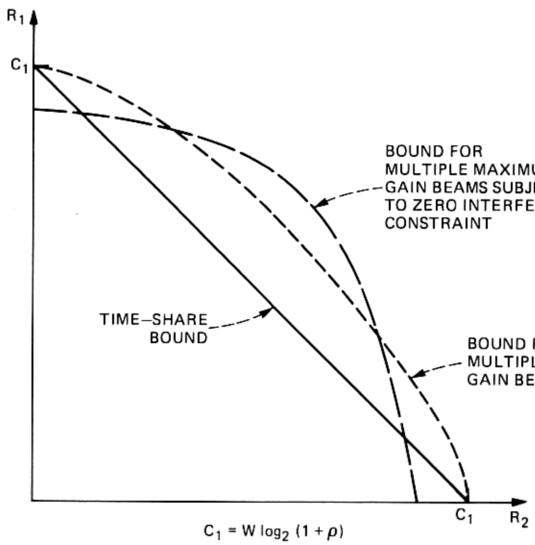

Fig. 6—Three typical capacity region inner bounds for two ground stations. ρ is the resulting carrier-to-noise ratio at a receiving ground terminal if all RF power available at the satellite is contained in a single beam of maximum gain in the direction of that ground terminal.

straints. Solution to this problem amply illustrates the relationship between ground-station separation and frequency-reuse advantages, while providing results that are easily presented and evaluated. Moreover, a study of the capacity region for two closely spaced ground stations provides, for the first time, a comparison of frequency-reuse techniques and simple time-sharing on the basis of available capacity. The analytical techniques used to evaluate the outer bound (32) are readily applied to study networks containing an arbitrary number of equally spaced ground terminals. A more general technique, described earlier in Section 3.1, can be applied to study systems in which the ground terminals are arbitrarily located.

To study the outer bound for a two-ground terminal network, we invoke the theory of discrete prolate spheroidal sequences (DPSS)[24,25] to find the supremum of (32). This theory is reviewed in Appendix A. Using the theory and notation contained therein, we now consider the expression for capacity (32). The G capacity-maximizing normalized radiation patterns $\{\psi_i\}$ are aperture-limited via (34) to $[-A, A]$. Invoking the sampling theorem, we can express $\psi_i(x)$ as

$$\psi_i(x) = X_o \sum_{n=-\infty}^{\infty} \psi_i(nX_o) \frac{\sin[(\pi/X_o)(x - n/X_o)]}{\pi(x - nX_o)}, \tag{58}$$

where

$$X_o = \frac{\pi}{A_o}, \qquad A_o > A. \tag{59}$$

In (58), we have again normalized the scale factor k/z_o to unity since, as will become apparent, the capacity region bounds are dependent upon the separation between ground stations relative to the null spacing of the narrowest beam which can be produced by the satellite aperture, rather than upon absolute distance.

Since $\int \psi_i^2(x)\, dx = 1$,

$$\sum_{n=-\infty}^{\infty} \psi_i^2(nX_o) = \frac{1}{X_o}. \tag{60}$$

Thus the normalization (33) implies the constraint (60) on the sample values of the radiation.

Let the G ground stations be located at X_o, $2X_o$, $\cdots$, GX_o. The expression for capacity (32) depends only on the vectors $[\psi_j(X_o), \psi_j(2X_o), \cdots \psi_j(GX_o)]$, $j = 1, \cdots, G$. Each of these vectors must therefore be linear combinations of the index-limited DPSS, or else a new set of vectors satisfying (36) could be found with greater energy at the location of the ground stations.

Restricting attention now to two ground stations, we can write

$$\psi_1(nX_o) = a\nu_1(nX_o) + b\nu_2(nX_o) \tag{61}$$

$$\psi_2(nX_o) = c\nu_1(nX_o) + d\nu_2(nX_o), \tag{62}$$

where $\nu_{1,2} = [\cdots\ \nu_{1,2}(-X_o),\ \nu_{1,2}(0),\ \nu_{1,2}(X_o),\ \nu_{1,2}(2X_o),\ \cdots]$ are proportional to the DPSS for $w_o = \pi(A/A_o)$, $N = 2$ (the proportionality constant is $(1/X_o)^{1/2}$). From (60) and (95), we find

$$a^2 + b^2 = 1 \tag{63}$$

$$c^2 + d^2 = 1. \tag{64}$$

Furthermore, from (96)

$$\sum_{n=1}^{2} \psi_1^2(nX_o) = \sum_{n=1}^{2} [a^2\nu_1^2(nX_o) + b^2\nu_2^2(nX_o)] = \frac{1}{x_o}[a^2\lambda_1 + b^2\lambda_2] \tag{65}$$

$$\sum_{n=1}^{2} \psi_2^2(nX_o) = \sum_{n=1}^{2} [c^2\nu_1^2(nX_o) + d^2\nu_2^2(nX_o)] = \frac{1}{X_o}[c^2\lambda_1 + d^2\lambda_2]. \tag{66}$$

Applying (36) and (96),

$$\alpha(1-\alpha)[\lambda_1 ac + \lambda_2 bd] = 0. \tag{67}$$

Thus, applying (38), (32) reduces to finding a, b, c, d and α such that

$$C_T^{(2)} = W\log[1 + g_1\alpha\rho] + W\log[1 + g_2(1-\alpha)\rho] \tag{68}$$

is maximized, subject to (63), (64), and (67), where

$$g_1 = \frac{a^2\lambda_1 + b^2\lambda_2}{D} \tag{69}$$

$$g_2 = \frac{c^2\lambda_1 + d^2\lambda_2}{D}. \tag{70}$$

In the above, $D = A/A_o \leq 1$ is the separation between ground station expressed as a fraction of the first null spacing of a single-mode radiation pattern possible with the aperture of extent $2A$. Figure 7 is the radiation pattern of a single maximum-gain mode directed at one ground station. This radiation possesses the familiar $(\sin X)/X$ pattern. For $D = 1$, the two ground stations are located as shown.

The eigenvalues λ_1 and λ_2 of (92) are dependent only upon D:

$$\lambda_1 = D + \frac{\sin \pi D}{\pi} \tag{71}$$

$$\lambda_2 = D - \frac{\sin \pi D}{\pi}. \tag{72}$$

For two ground stations, an outer bound for the capacity region is given by the intersection of the three regions

$$\mathscr{R}_1 + \mathscr{R}_2 \leq C_T^{(2)} \tag{73}$$

$$\mathscr{R}_{1,2} \leq W\log[1 + \rho]. \tag{74}$$

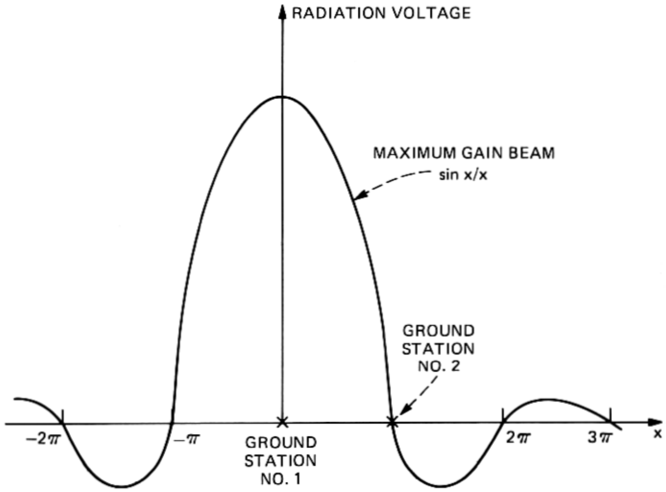

Fig. 7—Ground station locations defining the condition $D = 1$. By definition, D is the separation between ground stations expressed as a fraction of the beam center to first null of the narrowest beam that can be produced by the satellite antenna aperture.

Figure 8 shows the outer bounds for select ground station separations D and select carrier-to-noise ratios ρ (the rates have been normalized to b/s/Hz). In Fig. 9, we present additional bounds for lower values of ρ. For sufficiently high ρ, these curves clearly show that, as the distance between the two ground stations is reduced, the capacity region becomes smaller. At smaller ground-station separation distances, frequency reuse becomes less effective since, to maintain orthogonality of the modal received vectors, a smaller fraction of the energy of the second mode is directed toward the two ground stations. The water-fill solution then implies that, as the ground stations approach one another, a greater fraction of the available power is allocated to one mode to maximize the capacity.

For lower values of ρ, however, we see that the capacity region outer bound increases as D becomes smaller. This apparent anomaly is readily explained. For small ρ, there is insufficient power to excite more than one mode, and frequency reuse is of no advantage. As the distance between ground terminals is reduced, more power from this single mode is intercepted by the ground terminals, and the capacity region thereby increases.

Figure 10 plots the mode radiation patterns for a carrier-to-noise ratio $\rho = 9$ dB and a separation $D = 0.3$. In this case, 90 percent of the available energy is allocated to one mode. As the carrier-to-noise ratio increases, the two radiation modes remain the same but the power allocation changes such that, at $\rho = 21$ dB, both modes receive the same energy. Conversely, for values of ρ smaller than 9 dB, mode 2 receives no power at all.

Plotted in Fig. 11 as functions of separation D are the carrier-to-noise ratios below which the capacity is maximized by one-mode operation and above which the power allocation imbalance is less than 10 percent; that is, both modes receive approximately equal power. As D approaches unity, the CNR for each approaches $-\infty$.

Figure 12 plots an inner bound to the capacity region for select values of ρ and D. This bound is a composite of (a) the time share bound and (b) the bound obtained by exciting two noninterfering modes. For certain $(\mathscr{R}_1, \mathscr{R}_2)$ pairs, the greater inner bound is achieved by time-sharing; for other pairs, the opposite holds. For $G = 2$, we solve eq. (56) and normalize to find [see eq. (57)]:

$$g_1(X_o) = g_2(2X_o) = \frac{1 - (\sin \pi D/\pi D)^2}{\pi}. \tag{75}$$

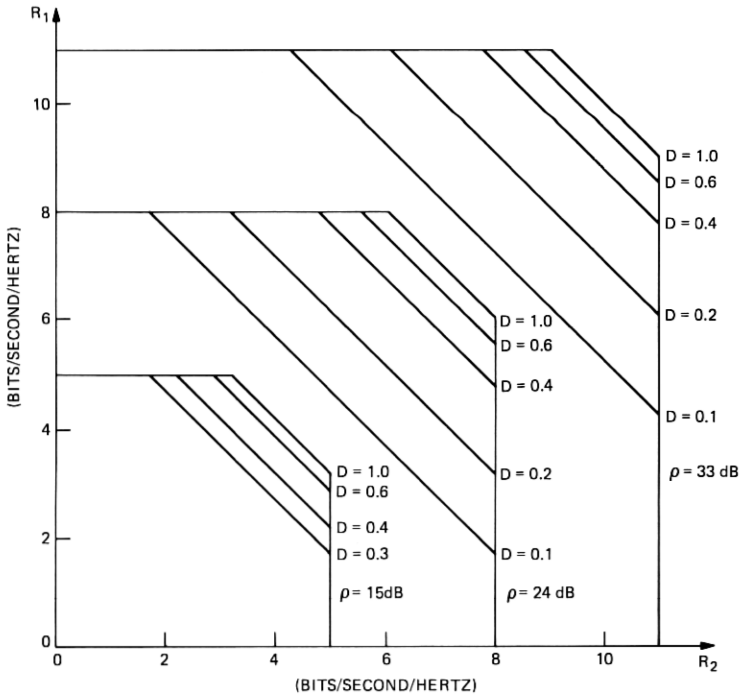

Fig. 8—Two ground-station, capacity-region outer bounds for select carrier-to-noise ratios ρ and select ground stations separations D. See captions for Figs. 6 and 7 for definitions of ρ and D.

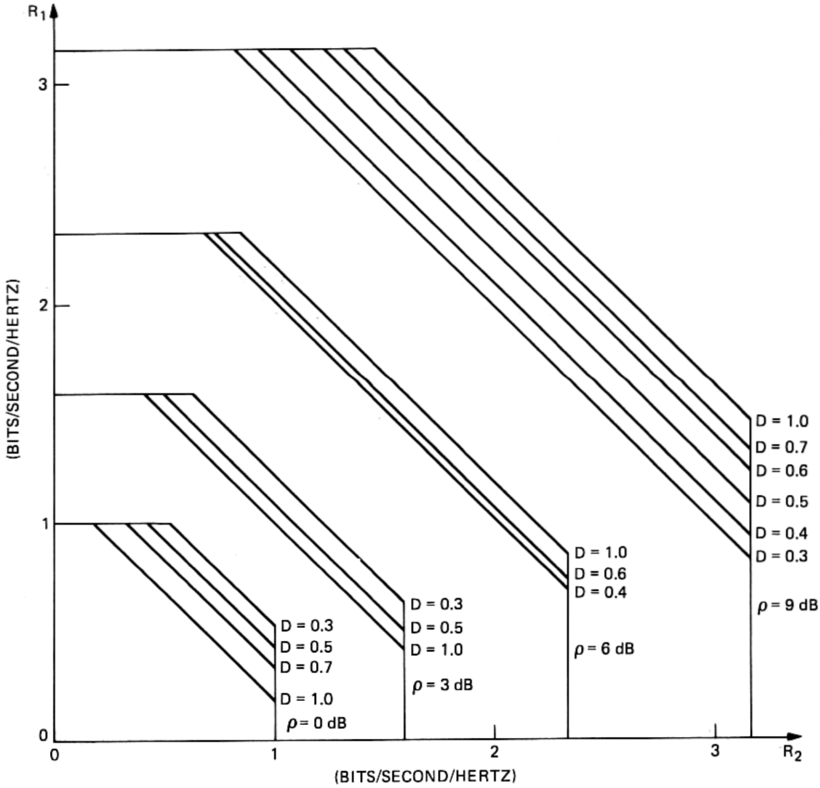

Fig. 9—Two ground-station, capacity-region outer bounds for select carrier-to-noise ratios ρ and select ground station separations D. See captions for Figs. 6 and 7 for definitions of ρ and D.

In Fig. 13, we consider the special case where the two ground stations require equal capacity ($\mathscr{R}_1 = \mathscr{R}_2 \hat{=} \mathscr{R}$). Two sets of curves are plotted. The first set shows, as a function of ground separation D, the difference in carrier-to-noise ratio between the upper bound and the noninterfering dual mode lower bound to maintain select transmission rates $\mathscr{R}$ into each ground station. From this curve can be inferred the maximum power saving possible via the most complicated coding and radiation schemes to achieve a rate $\mathscr{R}$, compared to interference cancellation. The second set of curves shows, as a function of separation D and for select values of $\mathscr{R}$, the power difference between time-sharing and interference cancellation to maintain that value of $\mathscr{R}$. For example, consider the case $\mathscr{R} = 3.5$ b/s/Hz, $D = 0.5$. To reliably maintain this transmission rate into each ground station via interference cancellation requires a carrier-to-noise ratio of 15 dB. The upper bound shows that, to maintain this rate, the minimum required CNR is 13.5 dB. Thus, the power saving possible by the most complicated coding and radiation scheme cannot exceed 1.5 dB. By contrast, for simple time-sharing, a CNR of 21 dB is needed. Thus, the method of interference cancellation achieves performance almost as good as the best theoretically possible and provides a power saving of 6 dB relative to time sharing. As shown in Fig. 14, for $D = 0.5$, both ground stations are well within the −3-dB contour of a single maximum-gain beam and therefore one might attempt to serve both either by the same fixed beam or by scanning a single beam between them; the results here show that it is possible to serve the two ground stations by two independent beams totally reusing the available spectrum and thereby save 6 dB in satellite power.

V. EXTENSIONS

The bounds on the capacity region for frequency reuse point-to-point communication satellites derived in Section II can be applied to other situations. We now discuss, qualitatively, two such extensions, confining attention exclusively to the outer bound.

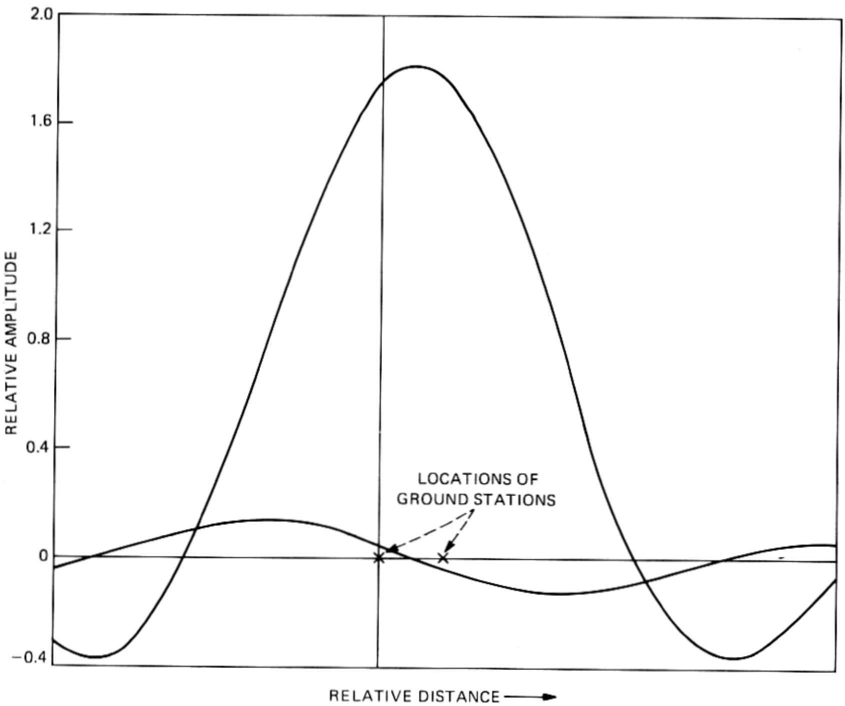

Fig. 10—Two mode radiation patterns for $\rho = 9$ dB and $D = 0.3$, weighted by optimum power allocation. See captions for Figs. 6 and 7 for definitions of ρ and D.

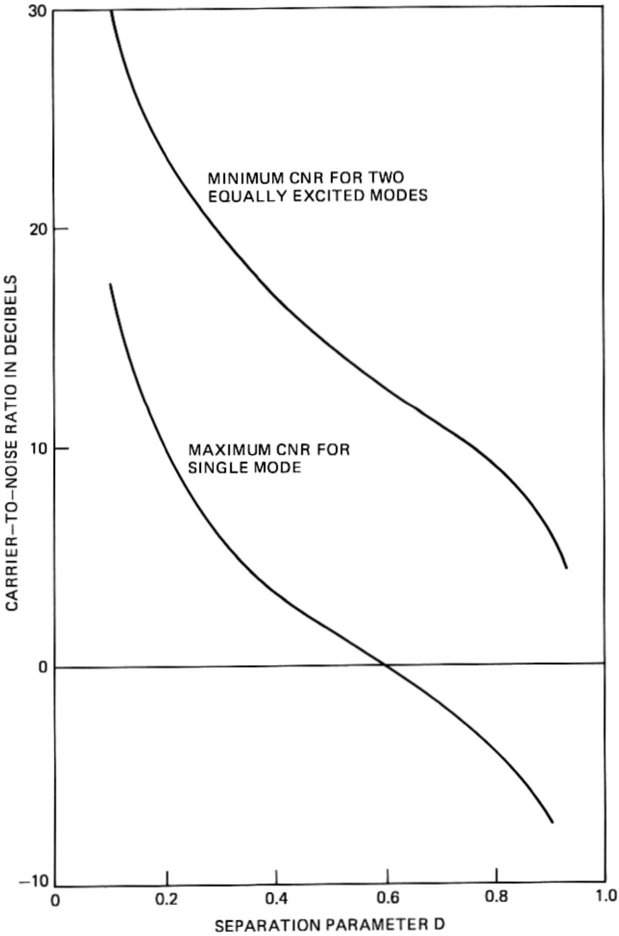

Fig. 11—Maximum CNR for single-mode operation and minimum CNR for equally excited dual mode operation vs ground-station separation.

At frequencies above 10 GHz, satellite communications are strongly influenced by precipitation.[26] The major effect of precipitation at these bands is the attenuation or fading of signal strength by heavy rainfall. The bounds of Section II were derived under the assumption that signal is available at all ground stations unattenuated. Thus these bounds always apply at frequencies below 10 GHz and apply for clear air conditions at higher frequencies. The bounds, however, can readily be extended to account for rain-induced fading.

Suppose that, due to rainfall, the signal voltage at the jth ground station is reduced to $\beta_j V$, where V is the clear air value and $0 \leq \beta \leq 1$. Typically, at any given time, $\beta_j = 1$ at most ground stations. For a particular set $\{\beta_j\}$, an upper bound to the capacity region for the G terminal network is then

$$C_T^G = \underset{\{\alpha\},\ \{\psi(x)\}}{\text{supremum}} \left\{ \sum_{j=1}^{G} W \log_2 \left[1 + \frac{\rho \alpha_j \sum_{k=1}^{G} \beta_k^2 \psi_{j,k}^2}{g_{\max}} \right] \right\}, \tag{76}$$

where

$$\alpha_j \alpha_l \sum_{k=1}^{G} \beta_k^2 \psi_{j,k} \psi_{l,k} = \alpha_j^2 \left(\sum_{k=1}^{G} \beta_k^2 \psi_{j,k}^2 \right) \delta_{j,l} \tag{77}$$

and conditions (33) through (35), (37), and (38) apply. Equation (77) is the orthogonality condition among modal received vectors, replacing (36), applying in the presence of rain fades.

The result (76) has an interesting interpretation. Suppose that, for clear air conditions, the vector $\underline{\mathscr{R}} = \{\mathscr{R}_1, \cdots, \mathscr{R}_G\}$ at which information

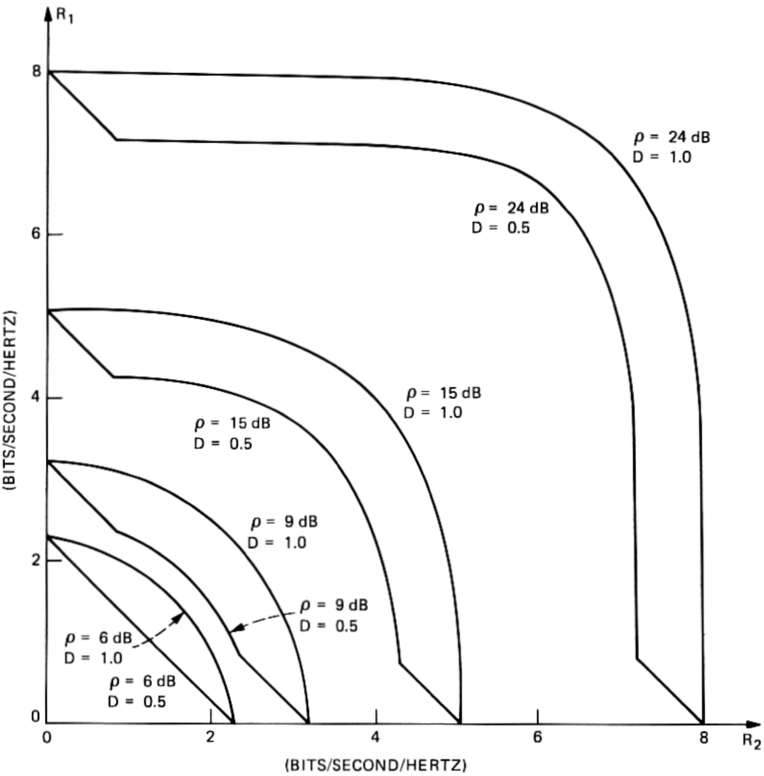

Fig. 12—Two ground-station, capacity-region inner bounds for various carrier-to-noise ratios ρ and ground station separations D. See captions for Figs. 6 and 7 for definitions of ρ and D.

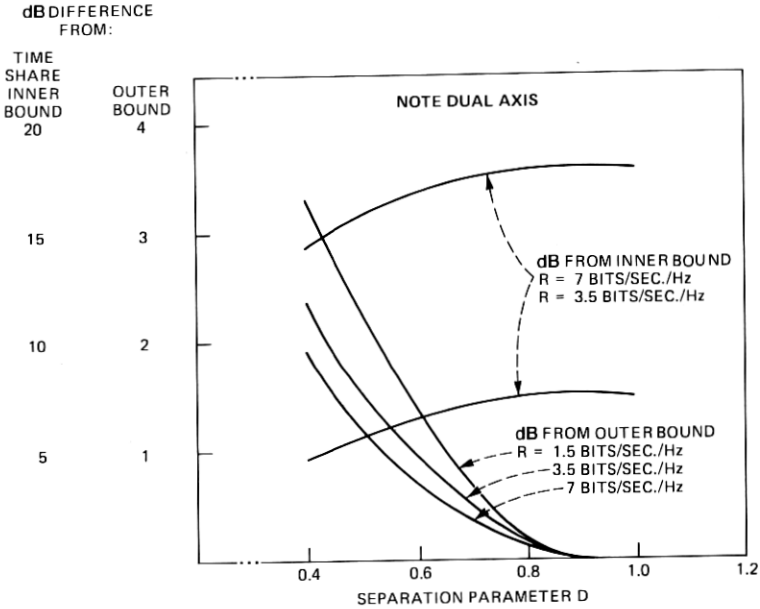

Fig. 13—Difference in decibels between the achievable capacity for two maximum gain, noninterfering modes and (a) the capacity region outer bound and (b) the time-share inner bound vs ground-station separation distance for selected equal rates into both ground stations. See Fig. 7 for definition of D.

is communicated to the ground stations is not on the boundary of the clear air capacity region. For various sets of fading conditions, corresponding to different attenuation factor sets $\{\beta\}$, we calculated the degraded capacity region via (76). Then, provided $\underline{\mathscr{R}}$ is within the degraded capacity region for all fading conditions of interest, we can reallocate the satellite power, bandwidth, and antenna resources such that reliable communication is maintained at the rates $\underline{\mathscr{R}}$. For example, to maintain the desired service availability, it may be necessary that up to F simultaneous fades, each of fade depth k dB, be accommodated. There are $G!/[F!(G-F)!]$ ways that the F fades can occur at the G ground stations. For each way, we find the resulting capacity region. Then, provided $\underline{\mathscr{R}}$ is within each capacity region, an allocation can be made to maintain service. For the typical case where $F \ll G$ and all ground stations require approximately the same information rate, the vector $\underline{\mathscr{R}}$ can be close to the boundary of the clear air capacity region. These observations are the essence of resource sharing approaches to combat rain fading, discussed elsewhere,[27] in that by accepting a traffic flow slightly smaller than the available clear air capacity, we can reallocate satellite resources to maintain service during fade events.

The capacity region outer bound developed in Section II can be applied to point-to-multipoint or broadcast satellites as well as point-to-point systems. For such an extension, we consider that, at the satellite, the G independent message sources, each one of which is intended for reception at a single ground station, are replaced by $2^G - 1$ independent message sources divided into G groups containing $G!/[k!(G-k)!]$ members, $k = 1, \cdots, G$ such that each source within the kth group generates information intended for reception at one combination of k ground stations out of the total of G ground stations. For example, there are three groups for $G = 3$ ground stations. Three sources M_1, M_2, and M_3 within the $k = 1$ group produce information at rates $\mathscr{R}_1$, $\mathscr{R}_2$, and $\mathscr{R}_3$ intended for reception at ground stations 1, 2, and 3, respectively. Three sources $M_{1,2}$, $M_{1,3}$, and $M_{2,3}$ within the $k = 2$ group produce information at rates $\mathscr{R}_{1,2}$, $\mathscr{R}_{1,3}$, and $\mathscr{R}_{2,3}$ intended for reception at ground stations 1 and 2, 1 and 3, and 2 and 3, respectively. One source $M_{1,2,3}$ within the $k = 3$ group produces information at the rate $\mathscr{R}_{1,2,3}$ intended for reception at all three ground stations. When deriving the outer bound, it was assumed that ground stations can communicate via a noiseless terrestrial link to a central processor. Thus, using (32), we can bound the rates of the seven broadcast satellite sources for $G = 3$ ground stations as follows:

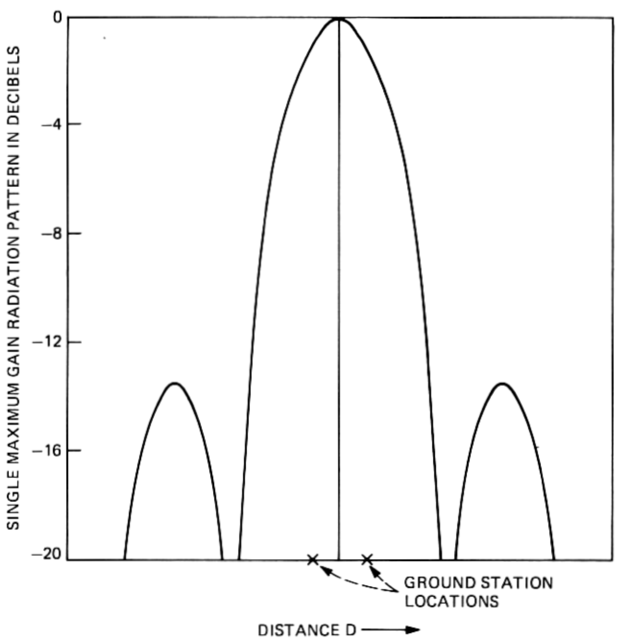

Fig. 14—Ground-station separation for $D = 0.5$ relative to a single maximum gain-beam radiation pattern. See caption for Fig. 7 for a formal definition of D.

$$\mathscr{R}_1 + \mathscr{R}_2 + \mathscr{R}_3 + \mathscr{R}_{1,2} + \mathscr{R}_{1,3} + \mathscr{R}_{2,3} + \mathscr{R}_{1,2,3} \leq C_T^{(3)} \tag{78}$$

$$\mathscr{R}_1 + \mathscr{R}_2 + \mathscr{R}_{1,2} + \mathscr{R}_{1,3} + \mathscr{R}_{2,3} + \mathscr{R}_{1,2,3} \leq C_T^{(2)}(1, 2) \tag{79}$$

$$\mathscr{R}_1 + \mathscr{R}_3 + \mathscr{R}_{1,2} + \mathscr{R}_{1,3} + \mathscr{R}_{2,3} + \mathscr{R}_{1,2,3} \leq C_T^{(2)}(1, 3) \tag{80}$$

$$\mathscr{R}_2 + \mathscr{R}_3 + \mathscr{R}_{1,2} + \mathscr{R}_{1,3} + \mathscr{R}_{2,3} + \mathscr{R}_{1,2,3} \leq C_T^{(2)}(2, 3) \tag{81}$$

$$\mathscr{R}_1 + \mathscr{R}_{1,2} + \mathscr{R}_{1,3} + \mathscr{R}_{1,2,3} \leq C_T^{(1)}(1) \tag{82}$$

$$\mathscr{R}_2 + \mathscr{R}_{1,2} + \mathscr{R}_{2,3} + \mathscr{R}_{1,2,3} \leq C_T^{(1)}(2) \tag{83}$$

$$\mathscr{R}_3 + \mathscr{R}_{1,3} + \mathscr{R}_{2,3} + \mathscr{R}_{1,2,3} \leq C_T^{(1)}(3). \tag{84}$$

where $C_T^{(2)}(i, j)$ is the total capacity possible into a network containing only ground stations i and j, $i = 1, 2, j = i + 1, \cdots, 3$, and $C_T^{(1)}(i)$ is the total capacity possible into a network containing only ground station i, $i = 1, 2, 3$. The intersection of regions (78) through (84) represent the capacity region upper bound for the sources of the broadcast satellite. We note that, if a particular ground station is included in the network for the purposes of calculating total capacity, then all sources communicating to that ground station must be included to establish the region. Generalization to any G is straightforward.

VI. SUMMARY AND CONCLUSION

In this paper, we have reviewed various concepts proposed to date for introducing high-capacity, spot-beam, frequency-reuse techniques for communication satellites. These proposals involve varying degrees of complexity in an attempt to achieve a uniform grade of service over a service region much wider than the beamwidth of a single spot beam which may offer nonuniform levels of traffic. The fundamental question unanswered by all these approaches is the ultimate capacity limit of a satellite constrained in power, bandwidth, and antenna aperture size.

Because the satellite system integrates multiple users, the ultimate capacity limit is expressed as a region over the space of transmission rates $\mathscr{R}_1, \cdots, \mathscr{R}_G$ into the G-ground terminals. Any particular set of rates contained in this region can be transmitted with arbitrarily high bit error-rate performance.

Outer and inner bounds for the capacity region were derived, and it was shown that the shape of the capacity region is dependent upon both the physical locations of the ground stations and the available carrier-to-noise ratio. Because of the Fourier transform relationship between the satellite aperture illumination and the far-field radiation pattern, the separation between ground stations, expressed as a fraction of the narrowest beam width which can be produced by the size-limited satellite aperture, affects the achievable degree of frequency reuse and therefore the capacity region. In the above, available carrier-to-noise ratio is defined to be the resulting CNR at a given ground station if all the radiated power available at the satellite is concentrated into the narrowest beam with maximum gain in the direction of that ground station.

To find an outer bound, a number of independent channels or modes equal to the number of ground stations were formed such that the components of a given mode as received at the ground stations were orthogonal to the components of all other modes. The various modes exhibit differing degrees of efficiency in directing power toward the ground stations. The total power of the satellite was then allocated to the modes in accordance with a classical water-filling problem; the modes exhibiting better directivity are given more power. The outer bound so obtained is all-inclusive, that is, no TDMA, FDMA, spread spectrum, antenna pattern shaping, or scanning spot beam approach can achieve a greater capacity region.

By definition, any scheme not requiring cooperating ground stations and for which the capacity region can be found provides an inner bound to the attainable capacity region since, by construction, that capacity region is actually achieved. The inner bounding schemes proposed included (*i*) time-sharing several maximum power-maximum directivity beams, (*ii*) formation of G independent but interfering beams, each with maximum possible gain in the direction of one ground station, and (*iii*) formation of G independent beams with the property that each beam exhibits maximum gain in the direction of one ground station subject to the constraint that no interference be produced at any other ground station. For a given CNR, the technique exhibiting the tightest bound was found to be dependent upon the rate set $\mathscr{R}_1, \cdots, \mathscr{R}_G$, that is, no one technique was found to provide the tightest bound over the entire capacity region. Thus, for these three schemes, the tightest inner bound is a composite of all three.

For uniformly spaced ground stations, it was shown that the theory of discrete prolate spheriodal sequences (DPSS) can be applied to facilitate calculation of the capacity region outer bound. The bounds were then applied to study, in detail, the effectiveness of frequency reuse for two closely spaced ground stations. A typical result shown there is that when the separation between the two ground stations is as small as 50 percent of the −3-dB beam width of the narrowest beam possible, it is desirable to employ frequency reuse via noninterfering beams.

Thus the capacity region bounds derived in this paper provide a yardstick against which the performance of any multiple spot-beam frequency-reuse technique can be compared. Research into antenna excitation techniques and multiple-user coding theory may reveal new concepts yielding inner bounds closer to the ultimate capacity region than those discussed here. Also, the need for new devices such as efficient linear power amplifiers, modulators, and power combiners is indicated to provide higher performance, ultimately approaching the capacity region outer bound.

Using the same approach as is presented here, it is possible to find outer and inner bounds applicable for satellites subject to more restrictive antenna constraints such as phased arrays. Of course, the capacity region so obtained must be smaller than that presented here, which was derived under the more general constraint that all aperture illumination functions vanishing outside the physical dimensions of the aperture are permitted.

APPENDIX

Discrete Prolate Spheroidal Sequences

Consider a sequence of real numbers a_n, $n = \cdots, -2, -1, 0, 1, 2, \cdots$, where $\sum_n a_n^2$ is finite. We define the spectrum of these numbers as

$$\mathscr{E}(\omega) = \sum_n a_n e^{j\omega n}. \tag{85}$$

Clearly,

$$\mathscr{E}(\omega + 2\pi) = \sum_n a_n e^{j(\omega+2\pi)n} = \mathscr{E}(\omega). \tag{86}$$

Moreover,

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathscr{E}(\omega) e^{j\omega n}\, d\omega. \tag{87}$$

We say that the sequence $\{a_n\}$ is bandlimited to $\omega_o < \pi$ if

$$\mathscr{E}(\omega) = 0, \qquad \omega_o \le |\omega| < \pi. \tag{88}$$

The energy in the sequence $\{a_n\}$ is defined to be

$$E = \sum_n a_n^2 \tag{89}$$

For a given index limit N and bandwidth ω_o, the DPSS of order 1 is defined to be that sequence $\{a_n\}$ for which

$$\lambda_1(N, \omega_o) = \frac{\sum_{n=1}^{N} a_n^2}{\sum_{n=-\infty}^{\infty} a_n^2} \tag{90}$$

is maximized, i.e., for a given index limit N, the DPSS of order 1 is that sequence bandlimited to ω_o which has maximum energy concentration over the index range [1, N]. Moreover, for index limit N, the DPSS of order $k \leq N$ is defined to be that sequence, bandlimited to ω_o and orthogonal to each of the DPSS of order 1, $\cdots$ $k-1$, such that

$$\lambda_k(N, \omega_o) = \frac{\sum_{n=1}^{N} a_n^2}{\sum_{n=-\infty}^{\infty} a_n^2} \tag{91}$$

is maximized.

The energy ratios $\lambda_k(N, \omega_o)$, $k = 1, \cdots, N$ are eigenvalues of the equation

$$\sum_{m=1}^{N} \gamma(n-m) a_{j,m} = \lambda_j a_{j,n} \tag{92}$$

where

$$\gamma(n) = \frac{\sin \omega_o n}{\pi n} \tag{93}$$

The index-limited DPSS are the eigenvectors of (92), i.e., the index-limited DPSS of order j is

$$\underline{a}_j = [a_{j,1}, \cdots, a_{j,N}]. \tag{94}$$

If we bandlimit the spectrum of each of (94) to ω_o and apply the transform operation (87), we obtain the DPSS. The DPSS are doubly orthogonal:

$$\langle \underline{a}_j \cdot \underline{a}_k \rangle = \sum_{n=-\infty}^{\infty} a_{j,n} a_{k,n} = \delta_{j,k} \tag{95}$$

$$\sum_{n=1}^{N} a_{j,n} a_{k,n} = \lambda_j \delta_{j,k}. \tag{96}$$

Consider now the sequence $\underline{b} = [b_1, \cdots, b_N]$. We wish to extend $\underline{b}$ outside the index range [1, N] such that the extended sequence is bandlimited to ω_o and has maximum energy concentration in [1, N]. The result is:[23]

$$\underline{b} = \sum_{j=1}^{N} c_j \underline{a}_j(N, \omega_o), \tag{97}$$

where $a_j(N, \omega_o)$ is the jth order DPSS and

$$c_j = \sum_{k=1}^{N} b_k a_{j,k} \tag{98}$$

REFERENCES

1. L. C. Tillotson, "A Model of a Domestic Satellite Communication System," B.S.T.J., *47*, No. 10 (December 1968).
2. W. G. Schmidt, "An On-Board Switched Multiple-Access System for Millimeter Wave Satellites," Proc. Int'l. Conf. Digital Sat. Comm., London, 1969.
3. R. Cooperman and W. G. Schmidt, "A Satellite Switched SDMA/TDMA System for Wideband Multibeam Satellites," ICC Conference Record, (June 1973).
4. T. Murantami, "Satellite-Switched Time-Domain Multiple-Access," EASCON, 1974.
5. G. Dill, T. Murantami, and R. Cooperman, "Application of SS-TDMA to Intelset Networks," Third Int'l. Conf. Digital Sat. Comm., Kyoto, Japan, 1975.
6. D. O. Reudink, "A Digital 11/14 GHz Multibeam Switched Satellite System," Sixth A1AA Comm. Sat. System Conf., Montreal, 1976.
7. Y. Ito et al., "Analysis of a Switch Matrix for an SS/TDMA System," Proc. IEEE, *65*, No. 3 (March 1977).
8. D. O. Reudink, A. S. Acampora, and Y. S. Yeh, "Spectral Reuse in 12 GHz Satellite Communication Systems," ICC Conf. Record, 1977.
9. H. W. Arnold, "An Efficient Digital Satellite Technique for Serving Users of Different Capacities," ICC Conf. Record, 1977.
10. D. O. Reudink and Y. S. Yeh, "A Scanning Spot Beam Satellite System," B.S.T.J., *56*, No. 8 (October 1977).
11. H. L. Van Trees, "Communication Satellites: Looking to the 1980's," IEEE Spectrum, *14* (December 1977).
12. P. F. Sielman, "The Role of Adaptive Multibeam Systems," Seventh A1AA Comm. Sat. System, Conf., San Diego, 1978.
13. A. S. Acampora, "Spectral Sharing in Hybrid Spot and Area Coverage Systems via Channel Coding Techniques," B.S.T.J., *57*, No. 7 (September 1978).
14. A. S. Acampora and B. R. Davis, "Efficient Utilization of Satellite Transponders via Time-Division Multibeam Scanning," B.S.T.J., *57*, No. 8 (October 1978).
15. D. O. Reudink and Y. S. Yeh, "The Organization and Synchronization of a Switched Spot Beam System," Fourth Int'l. Conf. on Digital Sat. Comm., Montreal, October, 1978.
16. H. S. Braham, "Characteristics of Several Multibeam Satellite System Applications for 1985 Operational Deployment," NTC Conference Record, December, 1978.
17. D. O. Reudink, A. S. Acampora, and Y. S. Yeh, "Methods for Achieving High Capacity Universal Service Satellites," NTC Conference Record, 1978.
18. T. M. Cover, "Broadcast Channels," IEEE Trans. Inform. Theory, *17-18* (January, 1972).
19. P. P. Bergmans, "Random Coding Theorems for Broadcast Channels With Degraded Components," IEEE Trans. Inform. Theory, *17-19* (March 1973).
20. A. B. Carleial, "Interference Channels," IEEE Trans. Inform. Theory, *17-24*, No. 1 (January 1978).
21. H. Sato, "An Outer Bound to the Capacity Region of Broadcast Channels," IEEE Trans. Inform. Theory, *17-24*, No. 3 (May 1978).
22. R. G. Gallagher, *Information Theory and Reliable Communication*, New York; Wiley, 1968, p. 344.
23. H. J. Landau, private communication.
24. D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—V: The Discrete Case," B.S.T.J., *57*, No. 5 (May–June 1978).
25. A. D. Wyner, "Signal Design for PAM Data Transmission to Minimize Excess Bandwidth," B.S.T.J., *57*, No. 9 (November 1978).
26. D. C. Hogg and T. S. Chu, "The Role of Rain in Satellite Communications," Proc. IEEE, *63* (September 1975).
27. A. S. Acampora, "A Shared Resource TDMA Approach to Increase in Rain Margin of 12/14 GHz Satellite Systems," B.S.T.J., *58*, No. 9 (November 1979).